

Implementasi Principal Komponen Analysis untuk Sistem Balik Citra Digital

Okfan Rizal Ferdiansyah.

Sistem Informasi, Fakultas Teknik Universitas Nusantara PGRI Kediri

Jl KH Achmad Dahlan no.76

okfanrizal.f@gmail.com

Abstrak – Kebutuhan citra digital meningkat dengan jumlah yang sangat besar untuk keperluan berbagai bidang seperti kedokteran, arsitektur, kesehatan, militer, olahraga dan berbagai bidang lainnya, seiring dengan bertambah besar penggunaan data citra digital diperlukan sebuah mekanisme untuk melakukan manajemen data tersebut, sampai saat ini banyak yang menggunakan mekanisme penotasian untuk proses manajemen citra digital, padahal dengan teknik ini akan menimbulkan perbedaan persepsi antara user dengan penotasinya, sehingga perlu di kembangkan sebuah sistem temu balik citra digital yang mampu mengatasi permasalahan tersebut. Besarnya dimensi citra menjadi sebuah masalah tersendiri bagi dunia komputasi, PCA mampu menjawab tantangan tersebut dengan melakukan proyeksi dari dimensi tinggi ke dimensi yang rendah. Pada penelitian sebelumnya PCA sudah sering digunakan untuk sistem temu balik citra digital hanya saja selalu menggunakan 100% komponen PCA, padahal terdapat potensi yang lebih untuk meningkatkan kecepatan dan akurasi sistem dengan menganalisa pengaruh jumlah komponen PCA yang digunakan dengan akurasi sistem. Pada penelitian ini peneliti mencoba melakukan analisa pengaruh jumlah komponen PCA yang digunakan terhadap akurasi sistem, selain itu juga dilakukan analisa performa PCA untuk mengetahui tingkat akurasi sistem dengan dataset University Washington, Visual Geometri Group dan dataset Wajah. Dari berbagai ujicoba yang dilakukan didapatkan hasil bahwa 10% s.d 30% komponen PCA yang digunakan menghasilkan akurasi sistem yang paling tinggi untuk masing-masing dataset, 91,4% untuk dataset University Washington 92,0% untuk dataset Visual Geometri Group dan 75,3% untuk dataset Wajah

Kata Kunci --Sistem temu balik citra digital, Citra digital, Principal komponen analisis.

I. PENDAHULUAN

Kebutuhan citra digital meningkat dengan jumlah yang sangat besar untuk keperluan berbagai bidang seperti kedokteran, arsitektur, kesehatan, olahraga dan bidang-bidang lainnya [1]. Besarnya jumlah citra digital membuat proses pencarian memerlukan waktu yang lama, sampai saat ini masih banyak yang menggunakan teknik notasi untuk melakukan proses pencarian citra digital[2], cara seperti ini sering terjadi kesalahan dalam proses pencarian sebuah citra karena proses penotasian sangat bergantung pada persepsi user.

Beberapa metode pernah diusulkan oleh para peneliti sebelumnya untuk proses sistem temu balik citra digital seperti *Sub Block Base Image Retrieval Modified Region Matching*[3], *Shape, Color and Relevance Feedback* [4]. *Sub Block Base* merupakan metode yang mampu melakukan reduksi dimensi dengan membagi sebuah citra digital menjadi beberapa blok region dan mengabaikan region-region yang tidak begitu memiliki informasi[3]. Akan tetapi metode ini memerlukan penelitian lebih lanjut untuk menentukan jumlah block region serta ukuran untuk mendapatkan hasil yang optimal. Metode *Shape, Color and Relevance Feedback* menggabungkan beberapa fitur untuk menghasilkan tingkat akurasi yang tinggi, tapi dengan jumlah data citra digital yang besar dapat mengakibatkan proses berjalan lambat, *Relevance feedback* juga sangat bergantung pada asumsi dari user [5].

Principal Component Analysis (PCA) adalah suatu analisis yang menjelaskan struktur varian-kovarian dari suatu himpunan variabel yang melalui beberapa kombinasi linear dari variabel-variabel tersebut [6]. PCA mampu

menjawab tantangan input sebuah citra digital dengan melakukan IMP proses reduksi dari dimensi tinggi ke dimensi yang lebih rendah. Penelitian sebelumnya tentang PCA untuk CBIR adalah Implementasi Pengenalan Wajah Menggunakan PCA [7], Implementation of CBIR System for CAD Jewellery Images Using PCA [8], dan CBIR Feture Vector Dimension Reduction with Eigenvectors of Covariance Matrix using Row, Column and Diagonal Mean Sequences [9]. Semua penelitian sebelumnya selalu menggunakan 100% komponen dari PCA, padahal terdapat potensi untuk lebih meningkatkan efisiensi waktu komputasi dengan menganalisa pengaruh prosentase komponen PCA yang digunakan. Sehingga pada penelitian ini peneliti fokus untuk menganalisa pengaruh reduksi komponen PCA terhadap akurasi sistem.

Penelitian sebelumnya tentang PCA dan CBIR pernah dilakukan pada tahun 2013 yang dilakukan oleh Kaur dan Jyoti yang berjudul Implementation of CBIR System for CAD Jewellery Images Using PCA[8], pada penelitian tersebut peneliti fokus pada presisi dan recall sistem. Dari hasil ujicoba didapatkan kesimpulan bahwa sistem mempunyai presisi sebesar 54%, recall 30% dan kecepatan waktu komputasi sistem bergantung pada query yang diinputkan oleh user.

PCA dan CBIR pernah diteliti oleh H.B. Kekre pada tahun 2010 berjudul CBIR Feature Vector Dimension Reduction with Eigenvectors of Covariance Matrix using Row, Column and Diagonal Mean Sequences [9]. Pada penelitian tersebut peneliti fokus pada analisa kombinasi matrix convarian yang digunakan terhadap performa dari PCA, didapatkan kesimpulan bahwa menggunakan matrix vector lebih baik dari pada menggunakan variasi vector didapatkan pula hasil dengan menggunakan citra digital true color mempunyai hasil yang lebih baik dibandingkan dengan citra grayscale.

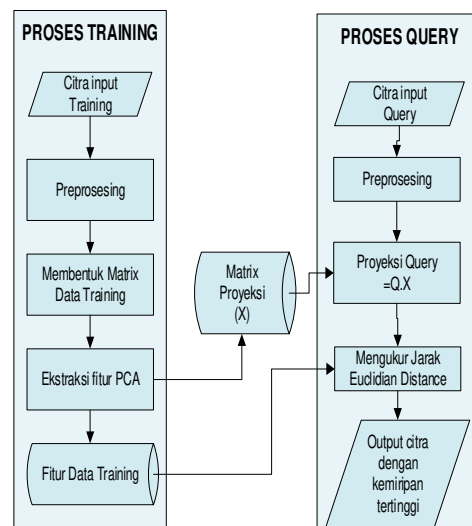
Selain itu PCA juga perna diteliti oleh Dian Esti Pratiwi dan Agus Harjoko pada tahun 2013 dengan judul Implementasi Pengenalan Wajah Menggunakan PCA mempunyai akurasi sebesar 82,81%[7]. Pada penelitian tersebut penulis ingin mengembangkan sebuah sistem identifikasi berdasarkan wajah, dari hasil ujicoba didapatkan kesimpulan bahwa banyak faktor yang mempengaruhi

akurasi sistem diantaranya adalah faktor cahaya, jarak dengan webcam, dan banyaknya gambar yang tersimpan di dalam database.

II. URAIAN PENELITIAN

A. Perancangan Sistem

Sistem yang dibangun adalah suatu prototipe yang digunakan untuk membantu peneliti dalam menganalisa CBIR, sistem memiliki beberapa attribute yang bisa diubah-ubah sesuai dengan kebutuhan peneliti untuk mendapatkan hasil yang diinginkan. Sistem yang dibangun mempunyai fungsionalitas yang terbatas yaitu untuk proses testing dan ujicoba, gambar 1 menunjukan flowchart sistem yang dibangun



Gambar 1. Flowchart Sistem

B. Principal Component Analysis

Dalam melakukan proses sistem temu balik dihadapkan dengan banyak variabel dengan data yang berdimensi tinggi. Operasi-operasi yang dilakukan terhadap citra query maupun citra yang ada di dalam dataset membutuhkan biaya komputasi yang sangat besar. Oleh karena itu, dibutuhkan PCA yang dapat mengurangi besarnya dimensi dari data yang di observasi yang signifikan dalam menggambarkan keseluruhan data.

C. Skenario Ujicoba

Beberapa skenario ujicoba dilakukan untuk mengetahui tingkat akurasi sistem, pada table 1 berikut ini adalah skenario-skenario yang

digunakan dalam sistem temu balik citra digital

Tabel 1. Skenario Ujicoba

No	Skenario	Training	Testing	Keterangan
1	Skenario I	50	75	Dataset UW
2	Skenario II	70	55	Dataset UW
3	Skenario III	90	35	Dataset UW
4	Skenario IV	40	60	Dataset VGG
5	Skenario V	60	40	Dataset VGG
6	Skenario VI	80	20	Dataset VGG
7	Skenario VII	30	120	Dataset Wajah
8	Skenario VIII	60	90	Dataset Wajah
9	Skenario IX	90	60	Dataset Wajah

Proses pertama yang dilakukan dalam implementasi sistem adalah memasukkan data *training* dan data *testing*, setelah itu sistem secara otomatis melakukan preprosesing, pembentukan data latih, ekstraksi fitur, proyeksi matrik query dan yang akan ditampilkan oleh sistem adalah hasil pengukuran jarak yang terdekat antara query dengan data training.

Selain itu setiap skenario akan dilakukan perubahan jumlah komponen PCA yang digunakan untuk mengetahui persentase reduksi dimensi yang paling baik dalam hal akurasi sistem seperti pada gambar 4

D. Analisa dan Ujicoba

Setelah dilakukan beberapa kali percobaan didapatkan hasil seperti pada tabel 4 berikut

Tabel 2. Hasil Ujicoba Training Vs Testing

No	Dataset	Training	Testing	Gambar	Akurasi
1	UW	50	75	Bunga	75,3%
				Bus	46,6%
				Dino	97,3%
				Gajah	82,6%
				Kuda	64,6%
				Rata-rata	73,3%
2	UW	70	55	Bunga	94,5%
				Bus	37,2%
				Dino	100%
				Gajah	86,3%
				Kuda	62,7%
				Rata-rata	76,1%
3	UW	90	35	Bunga	100%
				Bus	55,7%
				Dino	100%
				Gajah	98,5%
				Kuda	61,4%
				Rata-rata	83,1%
4	VGG	40	60	Pesawat	100%
				Daun	72%
				Motor	76%
				Gitar	96%
				Rata-rata	86%
				Pesawat	100%
5	VGG	60	40	Daun	91%

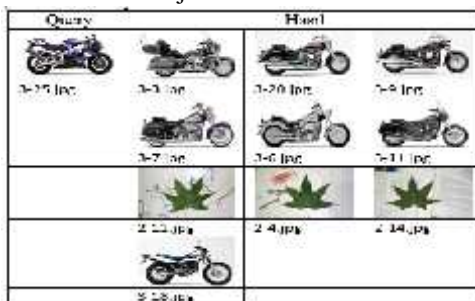
No	Dataset	Training	Testing	Gambar	Akurasi
6	VGG	80	20	Motor	73%
				Gitar	95%
				Rata-rata	89,7%
				Pesawat	100%
				Daun	88%
				Motor	68%
7	Wajah	30	120	Gitar	96%
				Rata-rata	88%
				Orang ke-1	24,5%
				Orang ke-2	47,0%
				Orang ke-3	49,0%
				Orang ke-4	46,0%
				Orang ke-5	45,5%
				Orang ke-6	45,0%
				Rata-rata	42,8%
8	Wajah	60	90	Orang ke-1	36,0%
				Orang ke-2	51,3%
				Orang ke-3	81,3%
				Orang ke-4	56,6%
				Orang ke-5	71,3%
				Orang ke-6	63,8%
				Rata-rata	63,8%
9	Wajah	90	60	Orang ke-1	46,0%
				Orang ke-2	53,0%
				Orang ke-3	92,0%
				Orang ke-4	99,0%
				Orang ke-5	71,0%
				Orang ke-6	79,0%
				Rata-rata	73,3%

Tabel 3. Hasil Ujicoba Nilai Komponen PCA

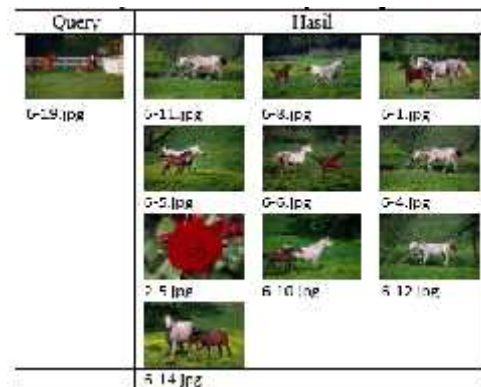
No	Dataset	Training	Testing	Komponen PCA	Akurasi
1	UW	50	75	100%	73,3%
2				90%	71,7%
3				80%	71,8%
4				70%	71,4%
5				60%	71,4%
6				50%	70,9%
7				40%	71,6%
8				30%	71,6%
9				20%	76,6%
10				10%	79,3%
11	UW	70	55	100%	76,1%
12				90%	75,8%
13				80%	75,6%
14				70%	74,5%
15				60%	73,8%
16				50%	73,4%
17				40%	74,7%
18				30%	75,0%
19				20%	77,6%
20				10%	78,7%
21	UW	90	35	100%	83,1%
22				90%	80,5%
23				80%	80,5%
24				70%	80,8%
25				60%	82,2%
26				50%	83,1%
27				40%	84,8%
28				30%	86,8%
29				20%	88,0%
30				10%	91,4%
31	VGG	40	60	100%	86,1%
32				90%	85,6%
33				80%	85,3%

No	Dataset	Training	Testing	Komponen PCA	Akurasi
34				70%	85,1%
35				60%	85,1%
36				50%	84,3%
37				40%	83,5%
38				30%	86,1%
39				20%	86,3%
40				10%	92,1%
41				100%	89,7%
42				90%	89,5%
43				80%	89,7%
44				70%	90,0%
45				60%	89,4%
46				50%	88,5%
47				40%	89,5%
48				30%	89,2%
49				20%	90,7%
50				10%	92,0%
51				100%	88,0%
52				90%	88,0%
53				80%	88,0%
54				70%	86,0%
55				60%	86,5%
56				50%	85,0%
57				40%	86,5%
58				30%	88,0%
59				20%	90,0%
60				10%	91,5%
61				100%	42,8%
62				90%	42,7%
63				80%	42,8%
64				70%	42,8%
65				60%	42,8%
66				50%	42,5%
67				40%	42,5%
68				30%	42,8%
69				20%	43,1%
70				10%	42,6%
71				100%	63,8%
72				90%	64,0%
73				80%	63,8%
74				70%	64,0%
75				60%	63,4%
76				50%	63,4%
77				40%	63,8%
78				30%	65,0%
79				20%	64,2%
80				10%	64,4%
81				100%	73,3%
82				90%	73,1%
83				80%	72,8%
84				70%	73,3%
85				60%	73,1%
86				50%	73,6%
87				40%	73,6%
88				30%	75,3%
89				20%	74,5%
90				10%	73,6%

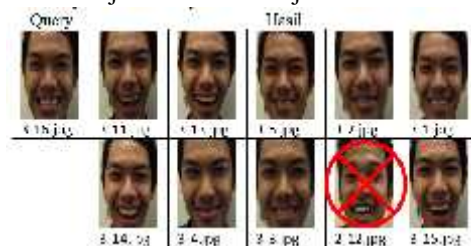
Tabel 4. Hasil Ujicoba dataset VGG



Tabel 5. Hasil Ujicoba dataset UW



Tabel 6. Ujicoba dataset wajah



III. SIMPULAN

- Akurasi rata-rata sistem yang paling baik untuk algoritma PCA dalam sistem temu balik citra digital dengan dataset *University Washington*, *Visual Geometri Grup*, dan dataset *Wajah* adalah sebagai berikut

 - Akurasi maksimal rata-rata sistem untuk dataset *University Washington* adalah 91,4% terjadi saat menggunakan 10% komponen dari PCA dengan 90 data *training* dan 35 data *testing*.
 - Akurasi maksimal rata-rata sistem untuk dataset *Visual Geometri Group* adalah 92,0% terjadi saat menggunakan 10% komponen dari PCA dengan 60 data *training* dan 40 data *testing*.
 - Akurasi maksimal rata-rata sistem untuk dataset *Visual Geometri Group* adalah 75,3% terjadi saat menggunakan 30% komponen dari PCA dengan 90 data *training* dan 60 data *testing*.
- Reduksi dimensi mempengaruhi tingkat akurasi sistem

 - Penggunaan antara 30% s.d 10% dari komponen PCA menghasilkan akurasi yang paling

- baik untuk semua jenis gambar dataset yang digunakan.
- b. Pilih 30% s.d 10% komponen mampu memangkas waktu komputasi hingga 90%.
 - c. Tingkat akurasi dengan reduksi dimensi tidak berbanding lurus maupun tidak berbanding terbalik, sehingga diperlukan penelitian lebih lanjut untuk menemukan pola reduksi dimensi dengan sebuah dataset.jumlah data training yang digunakan mempengaruhi tingkat akurasi sistem, semakin banyak data yang digunakan untuk proses training akurasi sistem cenderung semakin tinggi.

DAFTAR PUSTAKA

- [1] Shortliffe, E.H., Cimino, J.J., 2006, "Applications In Health Care And Biomedicine", Biomedical informatics:computer Springer.
- [2] Niket A, Kulkarni, R., 2013 Efficient Image Retrieval using Region Based Image Retrieval, International Journal of Applied Information Systems, ISSN : 2249-0868
- [3] Vimina E.R, and Jacop P.K, 2012, A Sub-bloc Based Image Retrieval Using Modified Integrated Region Matching, International Journal of Computer Science Issue, ISSN:686-692.
- [4] Yasmin, M., Mohsin, S., Irum, I., Sharif, M., 2013, Content Based Image Retrieval by Shape Color and Relevance Feedback, Life Science Journal, ISSN:1097-8135, 593-598
- [5] Gulhane, M.S., Shinde, A.G., Singh, A.K., 2013, Relevance Feedback for Content-Based Image Retrieval by Mining User Navigation Patters, Journal of Engineering, computer & Applied Science(JEC&AS), ISSN:2319-5606
- [6] Johnson, Richard. Dean Wichern. 2007. *Applied Multivariate Statistical Analysis*, 6th ed. New Jersey : Prentice Hall.
- [7] Pratiwi, D.E., 2013, Implementasi Pengenalan Wajah Menggunakan PCA (Principal Component Analysis), International Journal of Enterprise Information System, ISSN: 2088-37144.
- [8] Kaur, P., Jyoti, K., 2013, Implementation of CBIR System for CAD Jewellery Images Using PCA, International Journal of Scientific & Engineering Research Volume 4, ISSN:2229-5518.
- [9] Kekre, H.,B., 2010, CBIR Feature Vector Dimension Reduction with Eigenvector of Covariance Matrix using Row, Column and Diagonal Mean Sequences, International Jurnal of Computer Applications, ISSN:0975-8887.
- [10] Cahyadi, D., 2007. *Ekstraksi dan Kemiripan Mata pada Sistem Identifikasi Buron*. <http://lontar.ui.ac.id/file?file=digital/123280-SK-691> diakses tanggal 10 Mei 2014
- [11] Smith, L.I., 2002. A Tutorial on Principal Component Analysis. http://www.cs.otago.ac.nz/cosc453/student_tutorials/principal_component.pdf diakses tanggal 17 Mei 2014
- [12] Weisstein, E.W., 2007, Distance. From MathWord A Wolfram Web Resource <http://mathworld.wolfram.com/distance.html> diakses tanggal 17 Mei 2004
- [13] Manning, C.,D., 2009, An Introduction to Information Retrieval. Cambridge:Cambridge University Press.